CHAPTER 2
Principles of heterojunction photodiodes and their characteristics

2.1 Introduction

Most optical semiconductor devices contain at least one pn-junction. Understanding the physics of the pn-junction is, therefore, an important step in the analysis of semiconductor devices. This chapter will introduce the basic concepts of the pn-homojunction and semiconductor heterojunctions. The principles of PIN photodiodes and their characteristics will also be outlined in this chapter. These concepts will lead us in the way to understand their physical behavior and therefore, the heterojunction photodiodes can be designed in the next chapter.

2.2 pn-Homojunctions and semiconductor heterojunctions

Generally, the pn-junction in semiconductor is classified into two types: one is the homojunction and another is the heterojunction. The homojunction is formed by the same semiconductor material, while the heterojunction is formed with two dissimilar semiconductor materials, which can be either of the same doping type (isotype) or the opposite doping type (anisotype). Heterojunctions themselves are essential constituents of almost all electronic and optoelectronics devices. Therefore, it is very important to discuss these topics. However, the topic of both homojunctions and heterojunctions has been widely discussed, but here will, then, be limited to only the introduction of some basic concepts.

2.2.1 pn-Homojunctions

2.2.1.1 Electrostatics of the pn-homojunction: contact potential and space charge

It is important to realize what happens if n- and p-type semiconductors are brought in contact at room temperature. Electrostatics of the pn-homojunction in thermal equilibrium is shown schematically in Fig. 2.1. Under equilibrium conditions (when no external bias is applied), electrons in the n-type semiconductor, which $n = n_{NO}$, diffuse to the p-type semiconductor in which $n = n_{PO}$ and recombine with holes (majority carriers) in this region, while holes in the p-type semiconductor (p_{PO}) diffuse to the n-type semiconductor (p_{NO}) because of the concentration gradient. As holes continue to leave the p-side, some of the negative acceptor ions (N_A^-) near the junction are left uncompensated. Since the acceptors are fixed in the semiconductor lattice while the holes are mobile. Similarly, some of the positive donor ions (N_D^+) near the junction are left uncompensated as the electrons leave the n-side. There is therefore the formation of two regions (W_n and W_p), consisting of immobile ion of opposite types. The dividing line between the two regions, known as the metallurgical junction, occurs when $N_D - N_A = 0$. [16] The regions on both sides of the junction M consequently becomes depleted of free carriers in comparison with the bulk p and n regions. As a result, a negative space charge forms near the p-side of the junction and a positive space charge forms near the n-side. This space charge region, also known as the depletion region (W), creates an electric field that is direct from the positive charge (on the n-side) toward the negative charge (on the p-side). This electric field is called the built-in electric field. It has a maximum at the junction and goes to zero at the edge of the depletion region on both sides. By virtue of this electric field (ξ), a potential gradient develops, according to, [2]

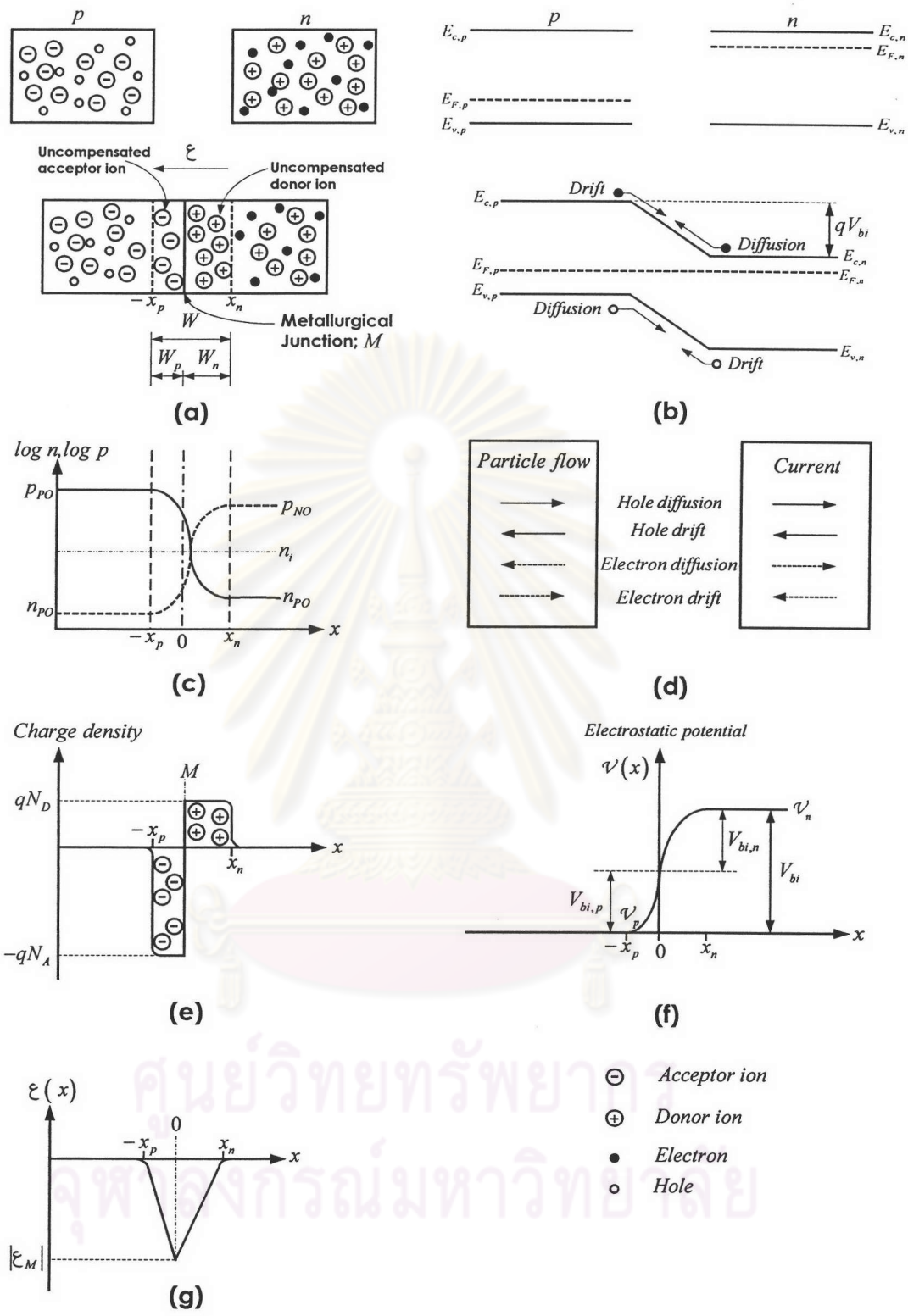


Fig. 2.1 Electrostatics of the pn-homojunction in thermal equilibrium
(a) Formation of pn-homojunction
(b) Junction band diagram
(c) The hole and electron concentration profiles
(d) Flow directions of the four-particle
(e) Space charge distribution
(f) Potential profile
(g) Electric field profile

$$V_{bi} = -\int \mathcal{E}\, dx \tag{2.1}$$

The bands bend in the depletion region by energy qV_{bi}, and this is called the contact potential or built-in potential. Looking at Fig. 2.1 (C), we are initially assuming an abrupt junction, in which a step change in impurity type occurs at the junction. Referring to the Gauss law, it follows that,

$$qN_A x_p = qN_D x_n = \epsilon_s \epsilon_0 \mathcal{E} \tag{2.2}$$

where ϵ_s is the static dielectric constant of the semiconductor and N_A and N_D are the doping densities on the p- and n-sides respectively. Also, full ionization of the dopants is assumed for this analysis. Equation (2.2) signifies that the total ionized negative charge per unit area on the p-side must be equal to the total ionized positive charge per unit area on the n-side, and this is known as the depletion approximation. From Poisson's equation for the abrupt junction,

$$-\frac{d^2V}{dx^2} = \frac{d\mathcal{E}}{dx} = \frac{q}{\epsilon_s \epsilon_0}\left[p(x) - n(x) + N_D^+(x) - N_A^-(x)\right] \tag{2.3}$$

Therefore,

$$-\frac{d^2V}{dx^2} \cong \frac{q}{\epsilon_s \epsilon_0} N_D, \quad 0 < x \le x_n \tag{2.4}$$

and

$$-\frac{d^2V}{dx^2} \cong \frac{q}{\epsilon_s \epsilon_0} N_A, \quad -x_p \le x < 0 \tag{2.5}$$

Upon integration, the electric field profile is given by

$$\begin{aligned} \mathcal{E}(x) &= \frac{qN_D}{\epsilon_s \epsilon_0}(x - x_n) \quad (n\text{-}side) \\ &= -\frac{qN_A}{\epsilon_s \epsilon_0}(x + x_p) \quad (p - side) \end{aligned} \tag{2.6}$$

Hence, the maximum value of the electric field is

$$\begin{aligned} \mathcal{E}_M &= -\frac{qN_D x_n}{\epsilon_s \epsilon_0} \quad (n - side) \\ &= -\frac{qN_A x_p}{\epsilon_s \epsilon_0} \quad (p - side) \end{aligned} \tag{2.7}$$

Integrating equations (2.4) and (2.5) twice gives

$$\begin{aligned} V(x) &= -\frac{qN_A}{2\epsilon_s\epsilon_0}(x_p + x)^2 \qquad (p-side) \\ &= -\frac{qN_D}{2\epsilon_s\epsilon_0}(x - x_n)^2 + V(x=0) \qquad (n-side) \end{aligned} \tag{2.8}$$

where $V(x=0)$ is the potential at the metallurgical junction:

$$V(x=0) = -\frac{qN_A x_p^2}{2\epsilon_s\epsilon_0} = -\frac{qN_D x_n^2}{2\epsilon_s\epsilon_0} \tag{2.9}$$

and it follows that

$$V_{bi} = \frac{q}{2\epsilon_s\epsilon_0}\left[N_A x_p^2 + N_D x_n^2\right] \tag{2.10}$$

2.2.1.2 The built-in voltage

At thermal equilibrium (without any external applied bias), the net current flow across the junction is zero. For this reason, the drift current due to the electric field must exactly cancel the diffusion current because of the concentration gradient. In symbols:

$$\begin{aligned} J_h(x) &= J_h(drift) + J_h(diffusion) = 0 \\ J_e(x) &= J_e(drift) + J_e(diffusion) = 0 \end{aligned} \tag{2.11}$$

Let us consider these current components in term of electric field and carrier concentration gradient

$$\begin{aligned} J_h(x) &= q\left(p\mu_h \xi(x) + D_h \frac{dp}{dx}\right) = 0 \\ J_e(x) &= q\left(n\mu_e \xi(x) + qD_e \frac{dn}{dx}\right) = 0 \end{aligned} \tag{2.12}$$

where carrier flow is assumed in the x-direction. For zero net hole current, $\xi(x) = \frac{D_h}{p\mu_h}\frac{dp}{dx}$ and remembering the equation (2.1), then we get

$$V_{bi} = \frac{D_h}{\mu_h}\int \frac{1}{p} dp \tag{2.13}$$

At any point x,

$$-V(x) = \frac{D_h}{\mu_h}\left|\ln p\right|_{p_{PO}}^{p(x)} \tag{2.14}$$

Thus,

$$V_{bi} = \frac{D_h}{\mu_h}\ln\frac{p_{PO}}{n_{NO}} \tag{2.15}$$

According to the Einstein relation, the ratio of D_h and μ_h is given by

$$\frac{D_h}{\mu_h} = \frac{kT}{q} \tag{2.16}$$

Using the Einstein relation, the equation (2.15) becomes,

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{p_{PO}}{n_{NO}}\right) \tag{2.17}$$

Similarly, starting with the equation for electrons,

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{n_{NO}}{n_{PO}}\right) \tag{2.18}$$

From which it follows,

$$n_{PO} p_{PO} = n_{NO} p_{NO} \tag{2.19}$$

The quantities in equation (2.19) represent equilibrium majority and minority carrier concentrations on both sides of the junction. Since in equilibrium the relation $n_o p_o = n_i^2$ holds, and assuming full ionization of the dopant impurity levels on either side (i.e., $p_{PO} = N_A$ and $n_{NO} = N_D$), one gets the familiar form,

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad \text{(V)} \tag{2.20}$$

In this equation V_{bi} is expressed in terms of known and measurable parameters of the materials forming the junction.

2.2.1.3 Depletion layer width

With respect to Fig. 2.1, in the space charge region where Poisson's equation is valid, the space charge density is equal to qN_D and qN_A on the n- and p-sides, respectively. Outside the depletion region, in neutral material on either side, the space charge density is zero. For an abrupt junction, the maximum electric field $\mathcal{E}_M$ is given by

$$\epsilon_s \epsilon_0 \mathcal{E}_M = qN_D x_n = qN_A x_p \tag{2.21}$$

The area enclosed by the electric field profile is the built-in voltage, V_{bi} which can therefore be expressed as

$$\left|-V_{bi}\right| = \frac{\mathcal{E}_M}{2}\left(x_n + x_p\right) \tag{2.22}$$

It follows that

$$W = x_n + x_p = \sqrt{\frac{2\epsilon_s\epsilon_0}{q}\left(\frac{1}{N_D}+\frac{1}{N_A}\right)\left|-V_{bi}\right|} \quad (2.23)$$

Hence, the depletion layer extends more into the side with the lower doping.

2.2.2 Semiconductor heterojunctions

As mentioned earlier, there are two types of heterojunction classified by doping type: anisotype and isotype, which can then be separated into four types: n-P, N-p, n-N and p-P, where the capital letter indicates the larger-bangap material. To analysis these heterojunctions, we need to built the energy band diagram according to Anderson's model on the assumption that the heterointerface is free of interfacial defect states. The energy band diagrams of isolated P-and n-type materials are shown in Fig. 2.2 (a) together with their vacuum level, work function; ϕ, electron affinity; χ, and Fermi level; E_F. With reference to the vacuum level, it is clear that the electron affinity of the wide-bandgap material is less than that of the narrow-bandgap material and therefore we can calculate the difference of both conduction band edge and valence band edge from its difference as in equation (2.24) to (2.26). The difference between the two conduction band edge energies is denoted by ΔE_c, and similarly, the difference between the two valence band edge energies is denoted by ΔE_v, [3]

$$\Delta E_c = \chi_1 - \chi_2 \quad (2.24)$$

$$\Delta E_v = \left(\chi_1 + E_{g1}\right) - \left(\chi_2 + E_{g2}\right) = \Delta E_g - \Delta\chi \quad (2.25)$$

Therefore,

$$\Delta E_g = \Delta E_c + \Delta E_v \quad (2.26)$$

After the two materials brought into contact (see Fig. 2.2 (b)), the Fermi energy levels are considered to be equalized and therefore the band bending occurs at both conduction band and valence band. The symbols qV_{b1} and qV_{b2} are assigned for the band bending, which can be determined from the Fermi level difference

$$E_{F1} - E_{F2} = \left(\chi_1 + E_{g1} - \delta_1\right) - \left(\chi_2 - \delta_2\right) = qV_{b1} + qV_{b2} \quad (2.27)$$

This energy difference corresponds to the total of the band bending, qV_{bi}. As in the case of the homojunction, the transition regions are assumed to be completely depleted over the distances x_1 and x_2, where, [11]

$$x_2 / x_1 = N_A / N_D \quad (2.28)$$

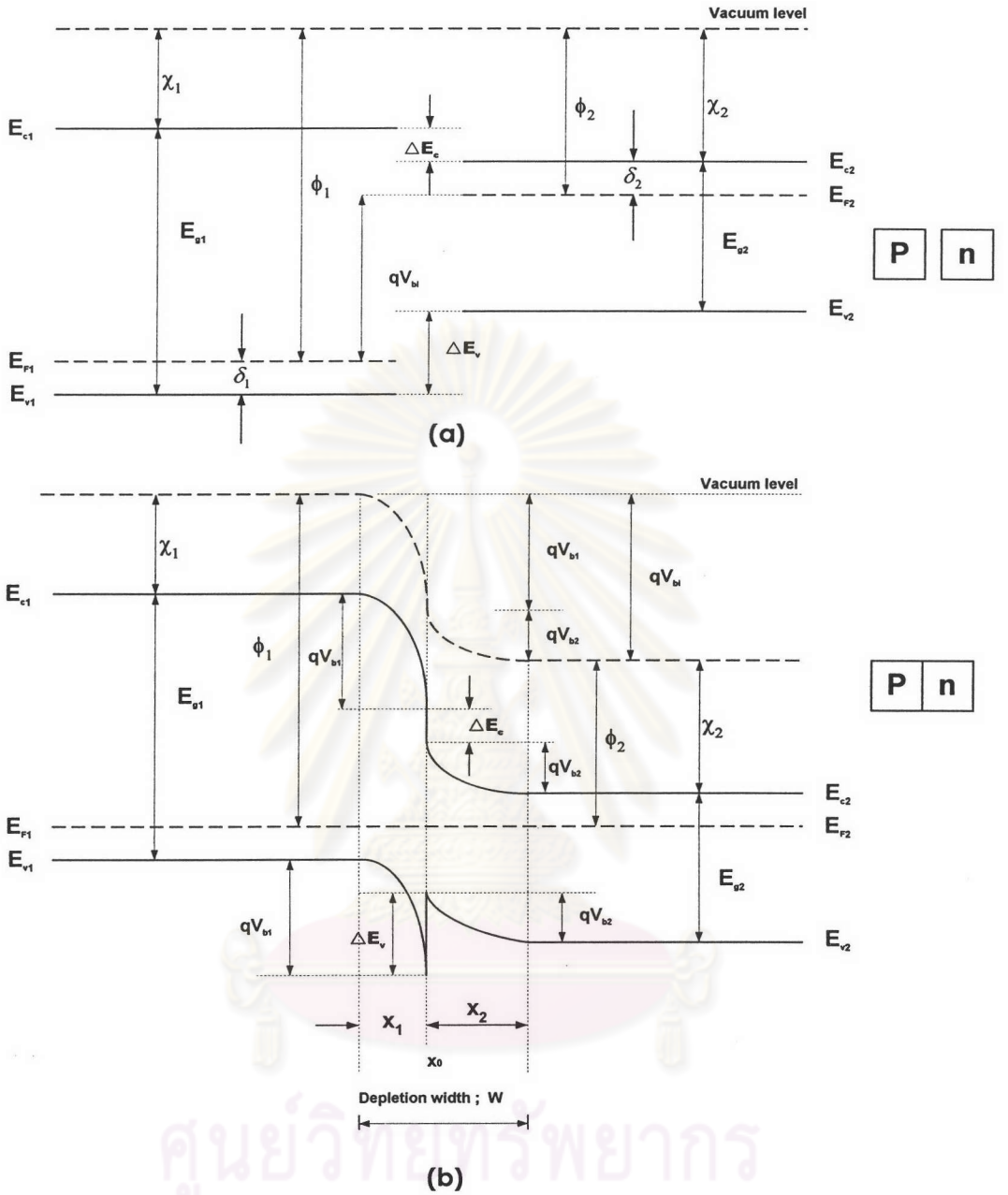


Fig. 2.2 Energy band diagram of an ideal P-n heterojunction (Anderson's Model) (a) before contact (b) at the thermal equilibrium

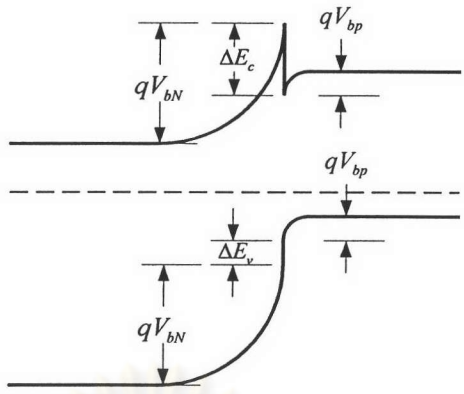
To satisfy the charge conservation, according to the Poisson's equation, we have

$$V_{b1} = \frac{\epsilon_2 N_D}{\epsilon_2 N_D + \epsilon_1 N_A} V_{bi} \tag{2.29}$$

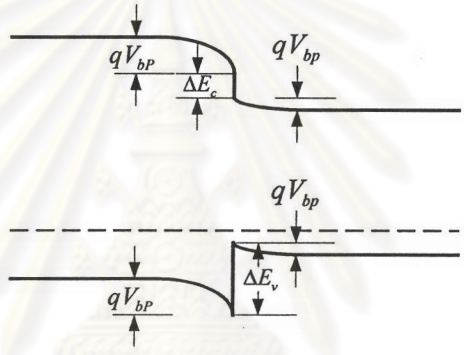
$$V_{b2} = \frac{\epsilon_1 N_A}{\epsilon_2 N_D + \epsilon_1 N_A} V_{bi} \tag{2.30}$$

Whence

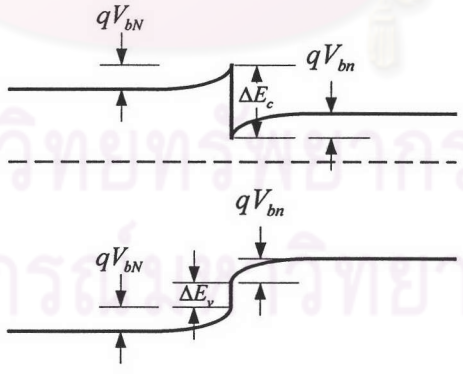
$$\frac{V_{b1}}{V_{b2}} = \frac{N_D \epsilon_2}{N_A \epsilon_1} \tag{2.31}$$



(a) Ideal energy band diagram of an N-p heterojunction in thermal equilibrium



(b) Ideal energy band diagram of a P-p heterojunction in thermal equilibrium



(c) Ideal energy band diagram of an N-n heterojunction in thermal equilibrium

Fig. 2.3 Ideal energy band diagram of the N-p, P-p and N-n heterojunctions

The corresponding depletion layer thickness in the p- and n-regions are given by

$$x_1 = \left[\frac{2 \epsilon_1 \epsilon_2 \epsilon_0 \, N_D V_{bi}}{q N_A \left(\epsilon_2 \, N_D + \epsilon_1 \, N_A \right)} \right]^{1/2}$$
$$x_2 = \left[\frac{2 \epsilon_1 \epsilon_2 \epsilon_0 \, N_A V_{bi}}{q N_D \left(\epsilon_2 \, N_D + \epsilon_1 \, N_A \right)} \right]^{1/2} \qquad (2.32)$$

The capacitance-voltage relationship is

$$C(V) = \left[\frac{q \, \epsilon_1 \epsilon_2 \epsilon_0 \, N_A N_D}{2 \left(\epsilon_1 \, N_A + \epsilon_2 \, N_D \right) \left(V_{bi} - V \right)} \right]^{1/2} \left(Farad.cm^{-2} \right) \qquad (2.33)$$

It is assumed here that the n- and p-regions are uncompensated and there is full ionization of the donor and acceptor centers in these regions. Note that these relations reduce to those for a homojunction when the dielectric constants (and bandgaps) of the two semiconductors become identical (i.e., $\epsilon_1 = \epsilon_2$). [2]

Similarly, the energy-band diagrams of the other types of heterojunction such as N-p, P-p and N-n are then considered and depicted in Fig. 2.3.

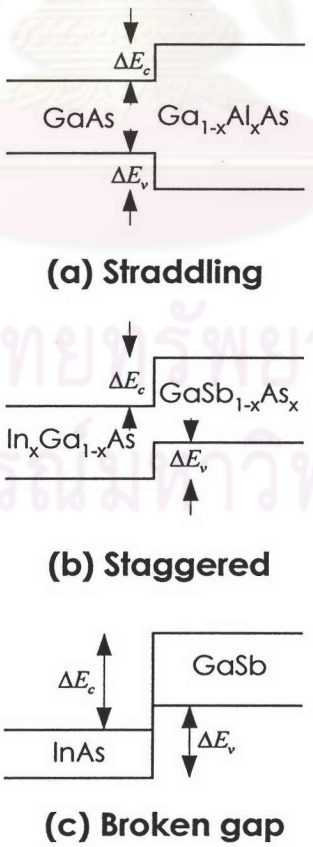


Fig. 2.4 Possibilities of band alignment in heterojunctions

As a consequence of two dissimilar semiconductors used to form a heterojunction, three types of band alignments can occur depending on their electron affinities and bandgaps. In the type we have considered so far, and shown in Fig. 2.4 (a), the valence band edge of the small bandgap is above the one of the large bandgap and therefore, the total difference in the energy gaps of the two semiconductors is shared by the discontinuities of $E_c(\Delta E_c)$ and $E_v(\Delta E_v)$, as mentioned in equation (2.26). The other two possibilities of band alignments are shown in Fig. 2.4 (b) and (c). In Fig. 2.4 (b), the valence band edge of the small bandgap is below the valence band edge of the large bandgap material, while in Fig. 2.4 (c) even the conduction band edge of the small bandgap semiconductor is below the valence band edge of the large bandgap semiconductor. These three alignment types are called straddling, staggered and broken gap respectively.

2.3 Principles of PIN photodiodes

2.3.1 Basic concept of photodiodes

Photodevices convert a light signal to an electrical signal such as a voltage or current. If the purpose of the photo-to-electrical conversion is to detect or determine information about the photo-energy, the device is called a photodetector. While the purpose is to produce electrical power, the device is called a solar cell. When the device is illuminated by light having energy greater than its bangap energy, the light is absorbed in the semiconductor and electron-hole pairs (Electron Hole Pairs: EHPs) are generated. The photogenerated carriers in the depletion region or within a diffusion length of it are then separated by the electrical field without any electric bias and an electromotive force between the p- and n-side semiconductors is generated. [4] As the carrier transverse the depletion region, a photocurrent is induced to flow out to the external circuit. The process of converting optical energy into electric energy is known as the photovoltaic effect. However, The solar cells use the photovoltaic effect without any electrical bias, while the photodiodes use the photovoltaic effect under an electrical bias. A bias circuit for pn-junction photodiode and current-voltage characteristics under light exposed are shown in Fig. 2.5.

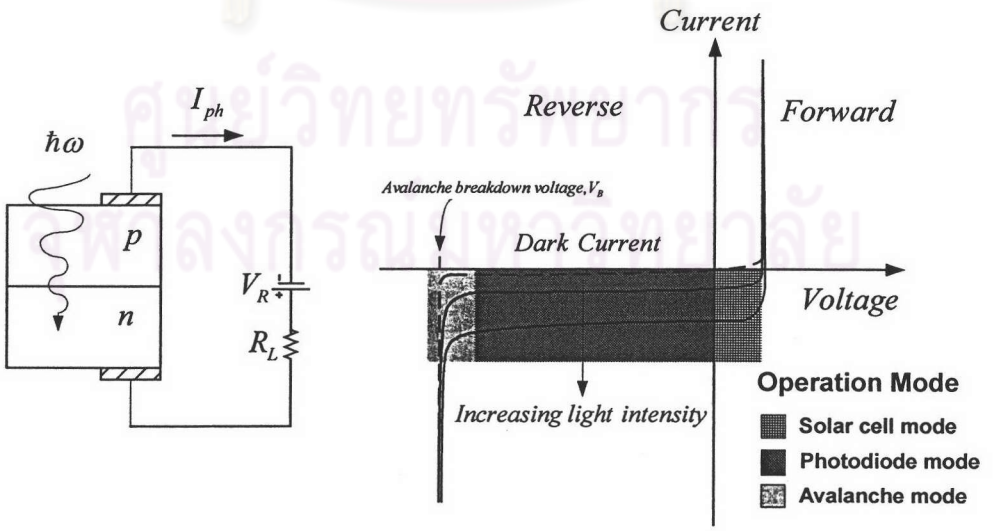


Fig. 2.5 The operation behaviors of pn-junction photodiode

The dash line indicates the current-voltage characteristics under no light illumination. The current under reverse bias without illumination light is termed the dark current. When increasing the light intensity, the characteristic will change as shown due to optical generation. As the result, the behaviors of photodiode can be classified into solar cell, photodiode and avalanche mode and particularly the photodiode and avalanche mode has generally three steps in the photodetection process [2]

1. Absorption of the optical energy and generation of carriers.
2. Transportation of the photogenerated carriers across the absorption and/or transit region, with or without gain (avalanche photodiodes have internal gain whereas p-i-n photodiodes have no internal gain but can have very large bandwidths).
3. Carrier collection and generation of a photocurrent, which flows through external circuitry.

2.3.2 PIN photodiodes and their characteristics

A p-i-n (or PIN) photodiode is a three-region structure in which an undoped region is ordinarily sandwiched between heavily doped p- and n-region. The i-region may consider to be either n^- or p^- region, which are respectively called the $\nu - layer$ and the $\pi - layer$. The characteristics of the p-i-n are strongly influenced by the thickness and doping concentration of i-layer. [4] Typically, the thickness of i-layer is much wider than p^+ and n^+ region. It is ordinarily about 5-50 μm depending on the particular application. [8] On the other hand, the carrier concentration is much smaller doping than those regions and is ordinarily between 10^{13} and 10^{16} cm^{-3}. [4] The p-i-n photodiodes have advantages over the p-n junction photodiodes in that they have a wider depletion region (thickness of i-layer), and therefore, there are a number of improvements. [8], [5]

- It increases the chances of an entering photon being absorbed because the volume of absorbent material is significantly increased.
- It reduces the junction capacitance. The lower the capacitance, the faster the device responses (lower RC time constant).
- Increasing the width of the depletion layer favors current carriage by the drift process, which is faster than the diffusion process.

Note that the thickness of i-layer should not be too wide, otherwise, it can lead to bandwidth degradation due to transit time effect and an increase in R of RC time constant. A schematic of the p-i-n photodiode operation under reverse bias is shown in Fig. 2.6. Since the p-i-n is in reverse bias, therefore in the dark, there is ideally no current flow at all. However, there is practically a small current cause by the random ionization of covalent bond within the depletion region. Heat causes the random breaking of a bond creating both a hole and a free electron. The free electron is attracted by the electric field towards the positive contact while hole is attracted towards the negative contact. This current is called the *dark current*, I_0 which is independent of the applied bias but varies with temperature. Thereby, the total current under illumination can be expressed as, [2]

$$I = I_{ph} + I_0 \qquad (2.34)$$

Where I_{ph} is the photocurrent produced by incident light. Under normal operating conditions, the photocurrent is several orders of magnitude larger than the dark current. The total current is therefore nearly equal to the photocurrent.

Because of the very low density of free carriers in the i-region and its high resistivity, any applied bias drops almost entirely across the i-layer, which is fully deplete. As the incoming photons are absorbed, the EHPs are then generated in the depletion region (i-region) and swept rapidly by the electric field to the opposite directions and give rise to the photocurrent. The photocurrent is composed of the *drift current*: J_{dr}, originated from the photo-induced carriers within the depletion region and the *diffusion current*: J_{diff}, due to the photo-induced minority carriers generated within their diffusion lengths from the edge of the depletion region. [21] The dark current is nearly equal to zero as mentioned earlier; therefore, we can simplify the total current density through the reverse-biased depletion layer to, [2]

$$J = J_{dr} + J_{diff} \tag{2.35}$$

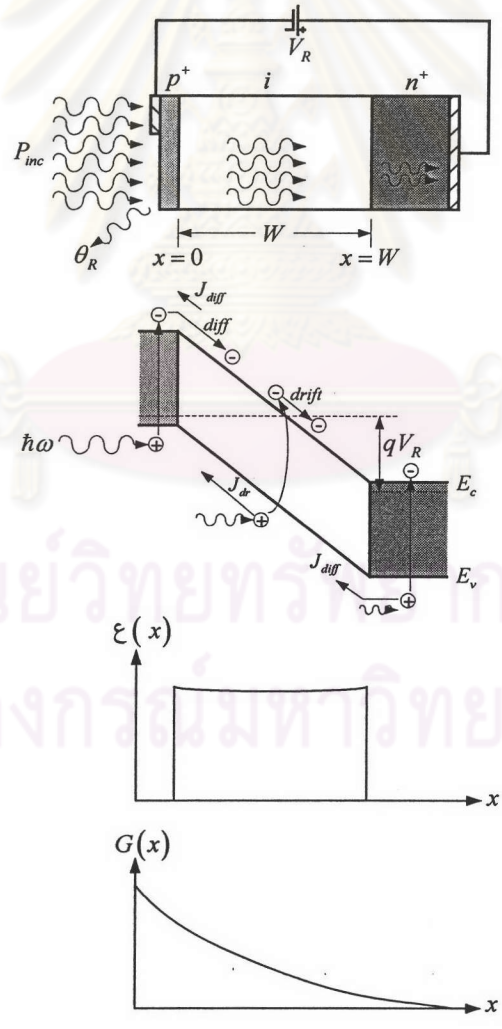


Fig. 2.6 A schematic of the p-i-n photodiode operation under reverse bias

Suppose that the thermal generation throughout the depletion region and the absorption in the top p^+ layer are negligible. The second condition implies that the thickness of the top layer is much smaller than α^{-1}, where α is the absorption coefficient of the incident photoexcitation. The origin of the distance coordinate is assumed to coincide with the p^+-i junction. The EHPs generation rate, assuming $\eta_i = 1$, is given by

$$\begin{aligned} G(x) &= \phi_0 \alpha e^{-\alpha x} \\ &= \frac{P_{inc}(1-\theta_R)}{Ah\nu} \alpha e^{-\alpha x} \end{aligned} \tag{2.36}$$

Where ϕ_0 is the incident photon flux (photons/sec/cm^2), θ_R is the reflectivity of the top surface, and A is the area of the diode, usually equal to the junction area. In practical diodes, the value of θ_R is made as small as possible by depositing an antireflection coating on the top surface. The drift current can then be expressed as

$$\begin{aligned} J_{dr} &= -q \int_0^W G(x) dx \\ &= -q\phi_0 \left(1 - e^{-\alpha W}\right) \end{aligned} \tag{2.37}$$

Where W is the width of the i-layer. Under the previous assumption, we are therefore concerned with the diffusion minority holes produced in the n^+ layer beyond the i-layer. The continuity equation for holes in this region at steady state, including generation and recombination, is given by

$$D_h \frac{\partial^2 p_N}{\partial x^2} - \frac{p_N - p_{NO}}{\tau_h} + G(x) = 0 \tag{2.38}$$

This equation can be solved with the boundary conditions

$$\begin{aligned} p_N &= p_{NO} \quad at\ x = \infty \\ p_N &= 0 \quad at\ x = W \end{aligned} \tag{2.39}$$

The solution of equation (2.38) is then of the form

$$p_N = p_{NO} - \left(p_{NO} + Ce^{-\alpha W}\right) e^{(W-x)/L_h} + Ce^{-\alpha x} \tag{2.40}$$

where

$$L_h = \sqrt{D_h \tau_h} \tag{2.41}$$

and

$$C = \frac{\phi_0 \alpha L_h^2}{D_h \left(1 - \alpha^2 L_h^2\right)} \tag{2.42}$$

The diffusion current is then given by

$$\begin{aligned} J_{diff} &= -qD_h\left(\frac{\partial p_N}{\partial x}\right)_{x=W} \\ &= -q\phi_0\frac{\alpha L_h}{1+\alpha L_h}e^{-\alpha W} - qp_{NO}\frac{D_h}{L_h} \end{aligned} \tag{2.43}$$

Adding equation (2.37) and (2.43), the total current density is given by

$$J = -q\phi_0\left(1-\frac{e^{-\alpha W}}{1+\alpha L_h}\right) - qp_{NO}\frac{D_h}{L_h} \tag{2.44}$$

Under normal operating conditions, the second term in equation (2.44) is small as the value of p_{NO} is normally very small. Hence, the total current is proportional to the incident flux ϕ_0, which is the expected result. The external quantum efficiency η_{ext} of the diode, then, can be express as

$$\begin{aligned} \eta_{ext} &= \frac{\left|J/q\right|}{P_{inc}/Ah\nu} \\ &= (1-\theta_R)\left(1-\frac{e^{-\alpha W}}{1+\alpha L_h}\right) \end{aligned} \tag{2.45}$$

For high quantum efficiency, a low reflection coefficient θ_R at the top surface is desirable. Another way, the term $\alpha W \geq 1$ is considered, which implies that α or W should be large. For a reason of transit time, W is limited, therefore, the high α should be desirable instead. However, the value of α cannot be changed too much, because it depends on the bandgap energy of semiconductor materials. If the diffusion coefficient L_h is very small, and $\eta_i < 1$, then the external quantum efficiency should be written as

$$\eta_{ext} = \eta_i(1-\theta_R)(1-e^{-\alpha W}) \tag{2.46}$$

Another important property of the photodiode is described by its responsivity, which is given by the current produced by the optical power incident [21].

$$\mathcal{R} = \frac{I_{ph}}{P_{inc}} = \frac{\eta_{ext}q}{h\nu} = \frac{\eta_{ext}\lambda(\mu m)}{1.24} \qquad A/W \tag{2.47}$$

Therefore, for a given external quantum efficiency, the responsivity increases linearly with wavelength as shown in Fig. 2.7.

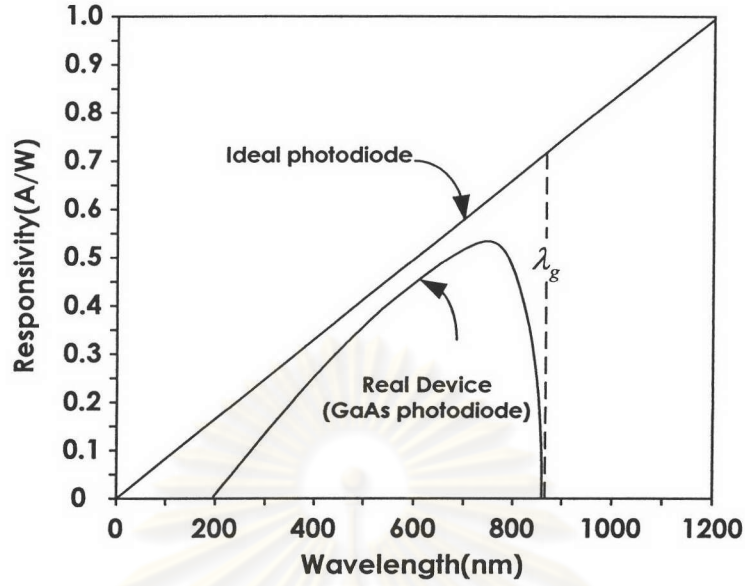


Fig. 2.7 A typically relation between wavelength (nm) and responsivity (A/W) at $\eta_{ext} \approx 1$

In addition to the quantum efficiency and the responsivity, the dynamic characteristic of a photodiode should be considered. Its performance depends on three limiting factors: (a) the time to diffuse of carriers created outside the depletion region within the diffusion length, (b) the time to drift across the depletion region, and (c) the time constant due to the junction capacitance originating from the depletion region.

Carriers generated outside the depletion region within a diffusion length can be diffused to the junction, but with a considerable time delay. To get rid of this diffusion effect, the generated carriers should be suppressed by the window effect or by the special design of the P^+ top layer. If the quantum efficiency is the primary interest, the depletion region should be sufficiently wide to absorb all of the incident light. However, the depletion layer must not be too wide, otherwise the transit time effects will limit the frequency response. On the other hand, making the depletion region too thin results in a large junction capacitance, which is given by

$$C = \frac{\in A}{W} \tag{2.48}$$

where A is the device area
W is the device depletion width
$\in$ is the dielectric constant of the employed material

The excessive diode capacitance, together with the load resistance, R_L, will result in a large time constant, $R_L C$. As mentioned above, for the purpose of high-speed response, very thin depletion layer is needed to reduce the transit time. However, this sacrifices the quantum efficiency. Thus there is a trade-off between the response speed and quantum efficiency.

2.4 GaAs-$Ga_{1-x}Al_xAs$ material system

GaAs is currently being used for a number of applications due to the direct bandgap itself and the high electron mobility. The cost is also the criteria to consider, as it has been widely used, therefore it is quite cheap. Furthermore, the technology has already been proven. The related compounds are GaAlAs, GaInAs, etc, but the most popular is GaAlAs. This is due to the lattice matching of the GaAs-$Ga_{1-x}Al_xAs$ heterostructure across nearly the entire Al composition range. The detail of the properties of GaAs-$Ga_{1-x}Al_xAs$ will discuss up next.

2.4.1 Energy bandgap

Since the $Ga_{1-x}Al_xAs$ is formed from the binary compounds, GaAs and AlAs, we will first consider the band structures of GaAs and AlAs shown in Fig. 2.8. There are three conduction band minima; direct Γ minimum, indirect X and L minimum of each compound. The conduction band of GaAs has its minimum at Γ so it is considered to be a direct bandgap material. In AlAs the direct Γ minimum is much higher than the indirect X minimum, so AlAs is an indirect bandgap material. In the $Ga_{1-x}Al_xAs$ alloy, compared to the GaAs, all of these conduction band minimum move up relative to the valence band as the composition x of Al increases. The energy variations of the three conduction band minima and the valence band maximum in $Ga_{1-x}Al_xAs$ as a function of the composition x are depicted in Fig. 2.9.

The lowest minimum in the conduction band changes from Γ (direct bandgap) to X (indirect bandgap) at $x \approx 0.45$. The change from a direct to an indirect energy gap gives a profound change in the optical properties. Usually $Ga_{1-x}Al_xAs$ is grown with the $x \leq 0.4$ to ensure that the lowest conduction band remains at Γ. The behavior for $x > 0.45$ is less well established. The bandgap of $Ga_{1-x}Al_xAs$ can be described as a function of the fraction x as, [2]

$$\begin{aligned} E_g^{\Gamma} &= 1.425 + 1.247x && \textit{direct - bandgap condition } x \leq 0.45 \\ &= 1.425 + 1.247x + 1.147(x - .045)^2 && \textit{indirect - bandgap condition } x \geq 0.45 \\ E_g^{X} &= 1.9 + 0.125x + 0.143x^2 \\ E_g^{L} &= 1.708 + 0.642x \end{aligned} \qquad (2.49)$$

The static dielectric constant ϵ_S and the electron affinity χ_e of the $Ga_{1-x}Al_xAs$ can also be written as a function of Al content (x) as, [1]

$$\begin{aligned} \epsilon_S &= 13.18 - 3.12x \\ \chi_e &= 4.07 - 1.1x \qquad \text{for } x \leq 0.45 \end{aligned} \qquad (2.50)$$

The variation of these parameters with Al composition allows us to realize the heterojunction, which is widely used in the design of optoelectronic devices.

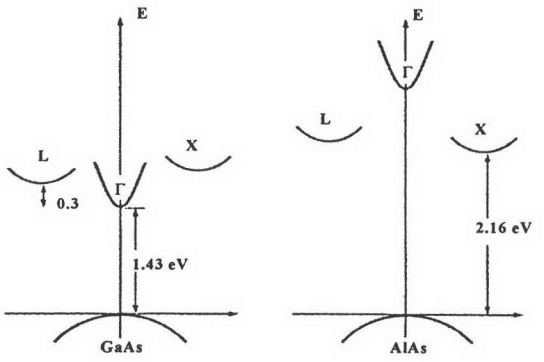


Fig. 2.8 Band structure of GaAs and AlAs

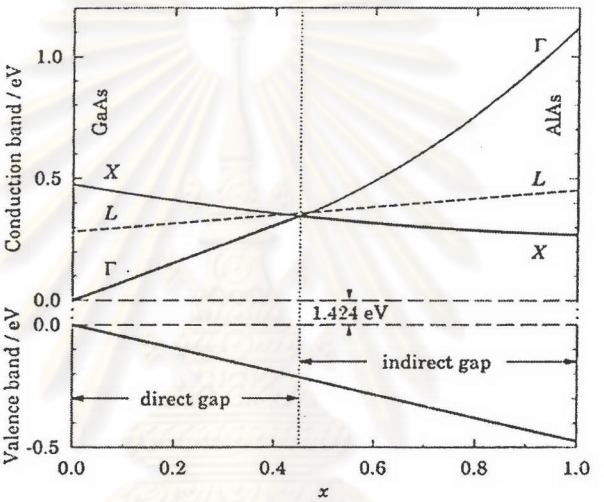


Fig. 2.9 Energy variations of the three conduction band minima and the valence band maximum in $Ga_{1-x}Al_xAs$ as a function of the composition x

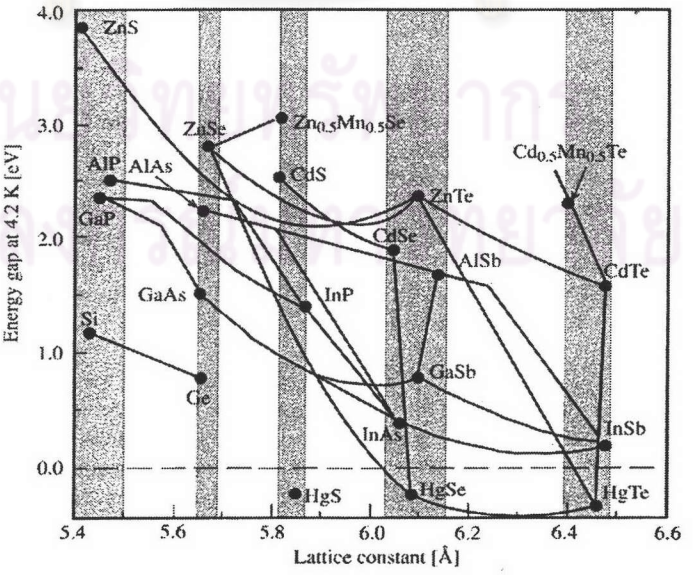


Fig. 2.10 Low temperature energy bandgaps of a number of semiconductors versus their lattice constants

2.4.2 Lattice constant

Lattice constant is also one of the parameters of the compound semiconductors depending upon the alloy composition. In Fig. 2.10, the low temperature energy bandgaps of a number of semiconductors are illustrated as a function of their lattice constants. The shaded vertical regions are drawn to indicate the groups of semiconductors with similar lattice constants or negligible mismatch. Materials within the same shaded region but having different bandgaps can, in principle at least, be combined to form heterojunctions with a particular band offset. The choice of band offsets can be widened by growing binary (such as SiGe), ternary (such as AlGaAs) and quarternary (such as GaInAsP) alloys indicated by the solid lines.

For example, the lattice constant (a) of $Ga_{1-x}Al_xAs$ as a function of the Al content (x) is given by

$$a = 5.6533 + 0.0078x \tag{2.51}$$

2.4.3 Absorption coefficient

The absorption coefficients for each semiconductor vary with the photon wavelength. For GaAs, the absorption coefficient is expressed as

$$\alpha = \begin{cases} 5.8\times10^4\sqrt{(E_p - E_g)}.....(E_a \le E_p) \\ \exp[100(E_p - E_g) + 7.82].....(E_b \le E_p \le E_a) \\ 0.13(E_g - E_p - 4.25\times10^{-2}).....(E_p \le E_b) \end{cases} \tag{2.52}$$

where E_p is the photon energy is given by hv/q (eV)

E_g is the bandgap energy, and

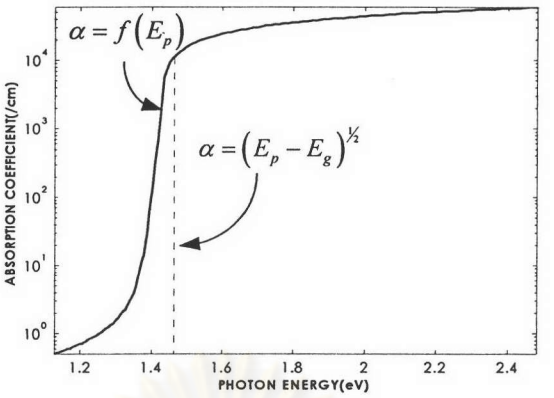
$E_a = E_g + 0.005\ eV$

$E_b = E_g - 0.0526\ eV$

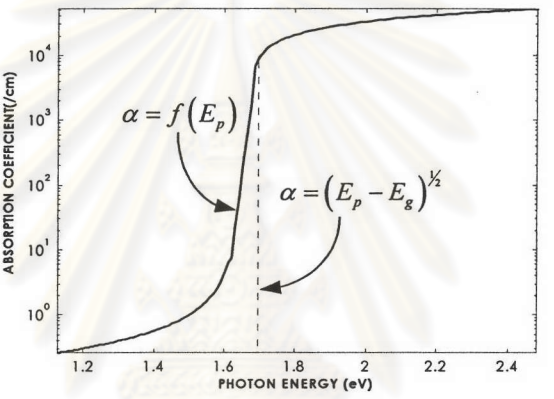
The absorption coefficient α of GaAs at room temperature, as shown in Fig. 2.11 (a), is theoretically calculated with equation (2.52). It should be pointed out that the values of the absorption coefficient change significantly near the band edge. This sharp rise of the absorption coefficient curve is known as the absorption edge. GaAs has a steep absorption edge because it is a direct gap material.

The energy bandgap of the alloy semiconductor is strongly dependent on the alloy composition. Therefore, the absorption coefficient is also dependent on the alloy composition. For example, in the case of GaAlAs alloy, the shape of the direct-gap absorption edge in GaAlAs alloy is quite similar to that in GaAs, as shown in Fig. 2.11 (b) and (c) for $Ga_{0.8}Al_{0.2}As$ and $Ga_{0.6}Al_{0.4}As$, respectively.

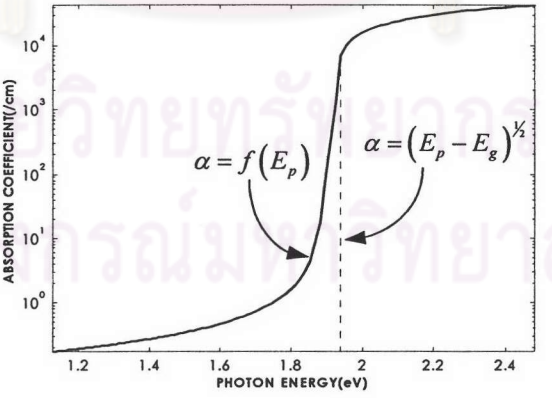
Since α is a strong function of the photon wavelength, for a given photodiode material, the wavelength range in which appreciable photocurrent can be generated is limited. Therefore, the absorption coefficient is one of the key factors that determine the quantum efficiency of a photodiode.



(a) Absorption edge of GaAs at room temperature



(b) Absorption edge of $Ga_{0.8}Al_{0.2}As$ at room temperature



(c) Absorption edge of $Ga_{0.6}Al_{0.4}As$ at room temperature

Fig. 2.11 Absorption edge of $Ga_{0.8}Al_{0.2}As$ and $Ga_{0.6}Al_{0.4}As$ at room temperature

2.5 GaAlAs/GaAs heterojunction and widow effect

Since, the heterojunction pair of $Ga_{1-x}Al_xAs/GaAs$ has a very small mismatch of the lattice constants, which is only 0.1% over the entire range of Al composition. For this result, there is only a few interface states existing at the junction between the two materials. Therefore, such a heterostructure can be grown free of a mechanical strain and significant imperfection.

When the $Ga_{1-x}Al_xAs$ (P^+)/GaAs (n^-) heterojunction is formed (see Fig. 2.12) and light is incident on a heterojunction from the $Ga_{1-x}Al_xAs$ side. Conceptually, $Ga_{1-x}Al_xAs$, which is a wide band-gap semiconductor would act as a window admitting photons of energy less than Eg1. Those of energy between Eg1 and Eg2 would pass freely through the $Ga_{1-x}Al_xAs$ and then absorbed in the GaAs. Thereby, the photogenerated carriers are created and this is known as the window effect.

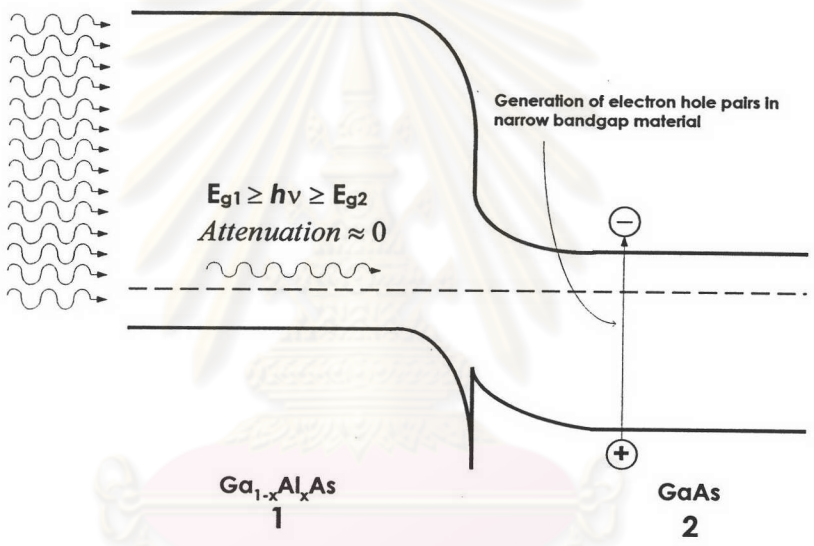


Fig. 2.12 Window effect demonstration on the $Ga_{1-x}Al_xAs$ (P^+)/GaAs (n^-) heterojunction

The principle advantage of window effect is reducing the surface recombination losses because it allows the junction to be placed deeper from the surface than in a homojunction. Therefore, in a heterojunction photodiode, the quantum efficiency does not critically depend on the distance of the junction from the surface. In addition, the heterojunction can provide unique material combinations so that the quantum efficiency and response speed can be optimized for a given optical-signal wavelength. For example, a $Ga_{0.6}Al_{0.4}As/GaAs$ heterojunction photodiode has a spectral response from ~650 to ~850 nm, which is wider than GaAs homojunction photodiode.

Anyway, we will account for the fact that as photons penetrate a material, their intensity decrease through absorption. In the same manner, when light enters the P^+-top layer, the fraction of light is absorbed. Hence, the power entering the depletion region is given by

$$P_{inc} = P_0 e^{-\alpha d} \tag{2.53}$$

where P_0 is the power incident on the P+ region (watt)

P_{inc} is the power entering the depletion region (watt)

α is the absorption coefficient (cm^{-1})

d is the width of the P+ region (cm).

Thus, the amount of optical power absorbed in the depletion region is given by

$$\begin{aligned} P_{absorb} &= P_{inc} - P_{inc} e^{-\alpha W} \\ &= P_{inc}\left(1 - e^{-\alpha W}\right) \end{aligned} \tag{2.54}$$

where W is the width of the depletion region

Therefore, with a smaller P+ region and a wider depletion region, the quantum efficiency of a photodiode increases.